

Visualizing the International Government Information Collection at University of Illinois in Urbana-Champaign

Uyen Nguyen

The University of Illinois at Urbana-Champaign (UIUC) has been a Federal Depository Library Program (FDLP) Library since 1907. Over the course of time, the Library has amassed one of the largest collections of government information with materials covering areas of agriculture, education, the environment, health, natural resources, and transportation.¹ In addition to federal and state publications in the United States, the Library also possesses an impressive collection of international government publications. The University Library became a United Nations depository in 1946 and Canadian depository in 1927, along with an extensive collection of British government resources.² The collection not only serves as a preservation of original documents from international agencies and governments for research, but also represents the diversity of the library collection and history at UIUC. While there have been many efforts to promote and focus on the federal information collection, the international government collection can be explored more to enhance visibility and usage of the materials.

This project stems from the desire to explore the international government collection at the University through data visualizations and compelling digital storytelling. The analysis will also present and illustrate the values of the collection within the context of other library services and collection at UIUC. These results can then provide an understanding of our materials and inform our decision to develop programs and shape government information services such as reference, instruction, and outreach efforts within the scholarly environment of the University and the public in the community.

This paper begins with a brief overview of data visualization and possible applications in a library context. Then, it will delve

to the methodology and share the findings of our research and analysis. It will also discuss limitations, challenges, and possible future directions to expand the project. Through these points, this paper aims to encapsulate the values of data in shaping a library's story and its collection, particularly focusing on the international government information collection.

Data Visualization in Libraries

Data visualization is the process of using computer-based systems to “provide visual representations of datasets designed to help people carry out tasks more efficiently.”³ Some examples of these visualizations range from common ones like bar or pie charts to more dynamic ones like word clouds or even animations. The choice of visualization can depend on the purpose of the project, quality of data data, and the skills of the users. Usually done at the end of the data lifecycle, data visualization allows better sharing of data and offers a more palatable and compelling ways to tell stories than just sharing raw numerical values. Additionally, at the core of any data-based projects should be the emphasis on human involvement and enhancement of our decision-making process. Graphical representation of data is not only useful to see trends and patterns but also offer creative ways to engage with the users digitally.

In libraries, data can come from various sources and be incredibly useful in helping library staff make decisions about library resources and future planning. For example, data can be used to understand collection development and financial budgeting.⁴ Libraries can employ data analysis on headcounts, reference interactions, or collection analysis. This information can help libraries make decisions regarding library resources

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
Row	Num of Lo	Governe	086 - Loca	Begin Pub	Material	Title	Language	MMS Id	Author	Publisher	Place of P	Subjects	Library Name	Location	C	Permaner	Holding Id	Physical D	Description
1	4	f	\$\$\$ GB\$	1918	Book	Report on the native	eng	99231085412205899	South-West Africa. A	Published United Kir	Namibia--	Main Stacks	stx	Q. 968	SO	22659838540005899	23659838530005899		
2	6	f	\$\$\$ (51	1988	Book	National policies a	eng	99147593012205899		OECD Pub France	Produce to	Oak Street Library	osx	338.10948	22665779230005899	23865779210005899			
3	0	f	\$\$\$ 1000-	1978	Book	Geology of Riggs an	eng	99629459912205899	Srivastava, P.	Ontario M Canada	Geology--I	Oak Street Library	osx	557.13	On	22638347850005899	236383478 no.174		
4	0	f	\$\$\$ 1000-	1978	Book	Geology of the Nort	eng	99629797412205899	Bennett, Gerald, 192	Ontario M Canada	Geology--I	Oak Street Library	osx	557.13	On	22708786470005899	237087864 no.163		
5	0	f	\$\$\$ 1000-	1978	Book	Geology of the Peter	eng	99629459312205899	Pyke, D. R.	Ontario M Canada	Geology--I	Oak Street Library	osx	557.13	On	22638277800005899	236382777 no.171		
6	0	f	\$\$\$ 1000-	1978	Book	Geology of the Horv	eng	99629459112205899	Breaks, F. W.	Ontario M Canada	Geology--I	Oak Street Library	osx	557.13	On	22638330710005899	236383306 no.169		
7	0	f	\$\$\$ 1000-	1978	Book	Ontario occurrence	eng	99567410812205899	Ferguson, Stewart A.	Ministry o Canada	Gold ores	Oak Street Library	osx	553	On8m	22798974600005899	237989745 no.17 (1978)		
8	0	f	\$\$\$ 10094	2005	Book	Qijiaochuan shi jia	chi	99532879312205899	Lin, Suzhen.		Taiwan--H	Main Stacks	stx	D8799.7	.L	22835693440005899	23835693430005899		
9	5	f	\$\$\$ 10095	2006	Book	Zheng zhi yu wen yi	chi	99629019612205899	Pu, Zhong-cheng.	Xing zhen China	Gao, Yi-sh	Main Stacks	stx	D8799.42	.T	22750971250005899	23750971240005899		
10	0	f	\$\$\$ 10096	2007	Book	Tai wan wai ji lao g	chi	99629032112205899		Zhong yar China	Lao gong--	Main Stacks	stx	HD8746	.T	22750935920005899	23750935910005899		
11	2	f	\$\$\$ 11-51:	1980	Book	Perspectives Canac	eng	9947653212205899		Available Canada	Canada--I	Main Stacks	stx	309.171	P	22720570760005899	23720570750005899		
12	0	f	\$\$\$ 11-52:	1987	Journal	Statistics Canada ir	eng	99704478912205899		Statistics Canada	Demograp	Oak Street Library	osx	CDROM 06	22830293320005899	238302933			2011
13	0	f	\$\$\$ 11-52:	1987	Journal	Statistics Canada ir	eng	99704478912205899		Statistics Canada	Demograp	Oak Street Library	osx	CDROM 06	22830293320005899	238302933			2005
14	0	f	\$\$\$ 11-53:	1999	Book	Electronic publicati	eng	99417974512205899	Ramage, Pamela.	Public Wo Canada	Depositor	Oak Street Library	osx	025.04097	22637871750005899	23637871740005899			
15	28	f	\$\$\$ 12-51:	1985	Book	Development and d	eng	9996637512205899	Platek, Richard.	Statistics Canada	Social sur	Main Stacks	stx	001.4	P696	22826185660005899	23826185650005899		
16	2	f	\$\$\$ 12-53:	1971	Book	Occupational class	eng	99299507512205899		Canada. Dominion B	Informatic	Main Stacks	stx	Q. 331.7	C	22726754950005899	23726754920005899		
17	0	f	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	2006:02:00	
18	1	f	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	2001:02:00	
19	1	f	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	2001:01:00	
20	1	f	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1981:01:00	
21	20	1	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1981:1A	
22	21	1	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1981:02:00	
23	22	1	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1986:02:00	
24	23	4	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1996:02:00	
25	24	1	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1991:01:00	
26	25	1	\$\$\$ 12-56:	1981	Journal	Standard geograph	eng	9911593812205899		Statistics Canada	Names, G	Main Stacks	stx	Q. 917.1	C	22763180730005899	237631806	1991:02:00	

Figure 1. A sample of our data exported from Alma

and opening hours, such as possibly extending the hours during finals.⁵ However, if a library only shows raw numerical data for a presentation, the audience may not be as engaged with the materials and may miss understanding important trends and key aspects. Popular data visualization tools like Excel and Tableau allow integration of both analysis and visualization, which can create compelling graphics for a wider audience. One can also pursue custom data visualization solutions, such as at Kingsborough Community College Library, which built SeeCollections, a visualization web application to represent their library holdings and features components of interactivity that allow for more dynamic interactions with the library's collection data.⁶ These examples show the endless possibilities of how data can be used and visualized in the management and future of libraries.

Combining the idea of “letting the data take the lead” with having pre-discussed questions, this project is exploratory and guided by collaborative discussion.⁷ Our results explore the stories that can be shaped by our data while providing solid guidance on how we want to present those stories. This project is also intended to be more of a top-down analysis of our collection to broadly understand our needs and resources.

Methodology

One of the most important steps in the process for creating data visualizations is the collecting and cleaning of data, which includes both defining the data types to include for the analysis process and the actual process of getting the data. For this project, international government collection covers publications produced by international government bodies and intergovernmental organizations and agencies such as the United Nations (UN), Food and Agriculture Organization (FAO), and World Bank. This broad approach ensured inclusion of a wide range of international government material.

For the data collection process, we extracted our data from Alma—the integrated library system (ILS) at UIUC. With Alma Analytics, a built-in analytics tool provided by our ILS, we built a report of the collection and exported them as csv files for the analysis process. Our report included 359,254 records and twenty-one criteria (columns) chosen for analysis, such as title and place of publication (e.g., country). These criteria were chosen based on consultations with the government information librarian, a literature review, and various questions we had about the collection:

- How big is the collection?
- Where are our materials coming from?
- What are the formats of the material?
- How frequently are the materials used?
- How has our collection grown over time?

Fortunately, the cleaning process was relatively quick because Alma was able to pull most of the necessary information. However, because Alma pulled the report based on the bibliographic records of the materials, there could be possible inconsistencies in some of these records because of workflows and policy changes over the years. The data cleaning process mainly involved spot checking these anomalies in the records. A majority of these records required manual corrections since there was no systematic way to clean this dataset. As an example, a scatter plot was created to represent the publication date of these publications, and one of the United Nations General Assembly documents (UN. A/HRC/35/22/Add.4) was dated 1017. This record was obviously inaccurate since the United Nations came into existence in 1945, so corrections were made in the dataset to this particular publication, as well as a few other titles that had similar problems.

Despite this challenge, this dataset was still relatively reliable as the main errors seemingly only concerned date-time data types, while the rest of the data types correctly corresponded to the bibliographic records.

While we initially used Alma Analytics and Alma's Data Visualization Tool for analysis, challenges arose due to a lack of documentation on some of the system functions. Excel and Tableau were considered as alternatives due to extensive scholarship and documentation related to data analysis. Ultimately, due to the size, variety, and purpose of our dataset and project, we decided to use Python and ArcGIS. Python was used for data analysis processing and creating graphical visualizations while ArcGIS StoryMap was used to present and share our findings with the broader community.

Python is currently one of the most popular programming languages used for data analysis, including data visualization, and is appropriate for a variety of purposes. For the project, we used common Python libraries for data-based projects such as Panda and Numpy for quantitative analysis, NLTK Toolkit for qualitative analysis, and Matplotlib for creating visualizations.

Besides these libraries, we also used a variety of other smaller and focused libraries for specific purposes in the process. For example, calendar is used to denote month and year data in the graphics. Despite Python's steep learning-curve, it was worthwhile to learn and explore its different capabilities for our collection data.

One important aspect of our project was to share our results with the broader community to highlight our collection and services, and for this purpose, we chose ArcGIS Storymap. Mainly used for creating cartographic narratives, Storymap is a useful tool to create interactive stories, and we were able to integrate the visualizations we created for this project. Furthermore, StoryMap allowed us to add context regarding our collection and its progression through time.

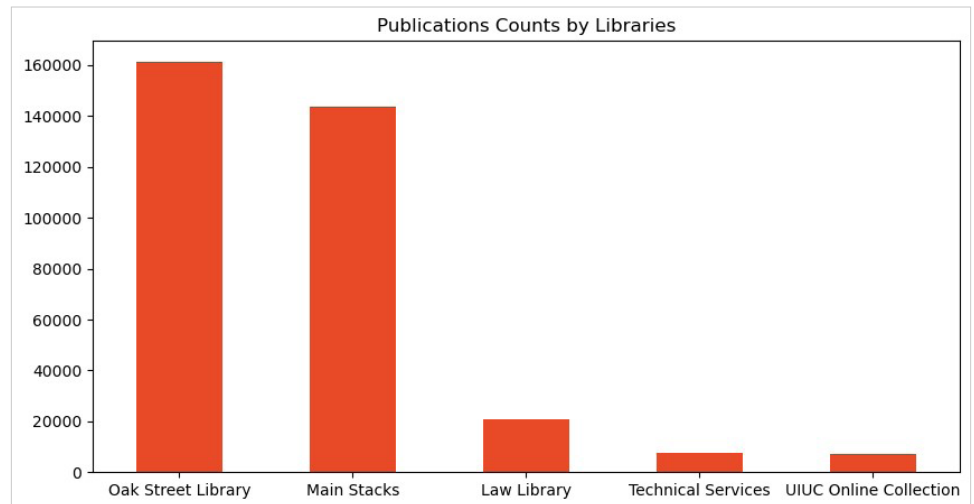


Figure 2. Number of publications by libraries

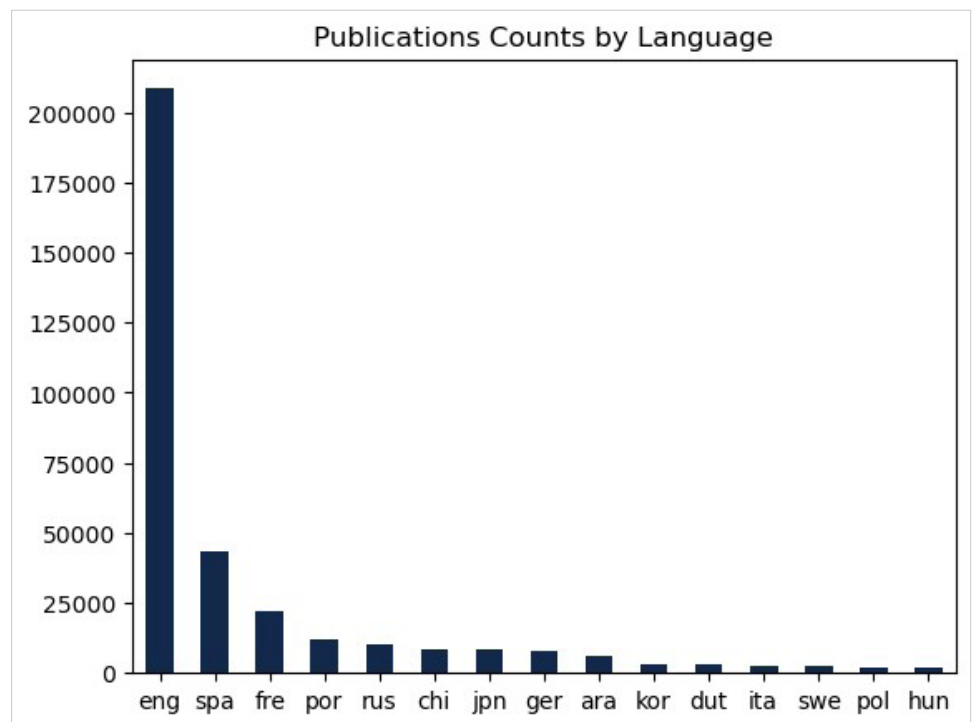


Figure 3. Number of publications by language

Results

The result of this project includes a series of visualizations created with Python and ArcGIS StoryMap, which will be published and shared with the community at UIUC in the near future. Below are some of the example graphics we included in our story.

Most of our collection is currently housed in the Oak Street Library and the Main Stacks. Oak Street Library is the high-density storage space of the University, and Main Stacks is the main storage space of the library collection at UIUC. Over the years, as we have received more publications and

materials and shifted library spaces and organizational structures, the collection has become dispersed across different units. However, these two locations remain the main storage spaces at UIUC, which helps with access and preservation purposes for our users and library staff to manage and maintain our collection.

While most of the collection is in English, we also have materials in a variety of other languages. This is essential for us to understand the diversity of our collection and gaps for future collection development. Unlike traditional collections in the library where the subject selectors can make intentional choices to shape their collection, a lot of the materials in the government information collection consists of donations and purchases from different international government bodies and agencies. Because of this, the content may not always align with what the Library has in mind for collection development. This requires careful consideration in developing the collection and deciding what to display for the wider community. In our collection at UIUC, we also have materials from less common languages such as Galician, Nzima, and Soninke, which we hope to highlight through this project.

Figure 4 depicts the collection counts by year of publication, representing both the growth of the international government information collection at UIUC and a reflection of the world's progression in publishing. For example, the United Nations was founded in 1945 and because UIUC became a UN depository in 1947, we have received a huge number of publications from the UN since then. While we don't necessarily receive all UN publications that are available, this graph demonstrates an increase in knowledge production across the globe.

Besides creating graphic visualizations, Python enabled us to identify major themes in our collection, such as subjects and countries of origin. While these analyses may not always result in graphically appealing images, they provide useful insights into our collection, helping us create a more comprehensive story of our collection.

Limitations and Challenges

While this project provided a useful overview and understanding of our international documents collection, there are limitations in some areas. Firstly, we have not explored qualitative data analysis in depth, which could offer a more comprehensive

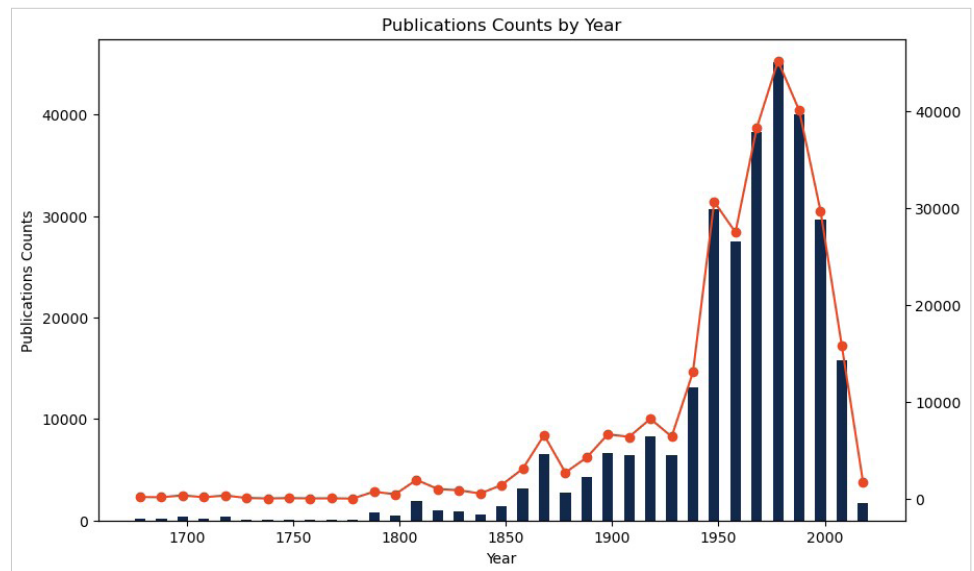


Figure 4. Publications counts by Year of Publication

picture of these publications. Our project mainly focused on quantitative data to get a broad view of this collection, thus we did not have the chance to deep dive into more nuanced elements. This limitation means our understanding of the collection may not capture the full range of factors and contexts that influence its use and value to users. Secondly, there was not a systematic workflow for cleaning our data. Consequently, there may be some anomalies that weren't fully addressed in our datasets. While we have confidence in Alma and its collection reports, this is a weakness we want to acknowledge.

As this project progressed, we also encountered various challenges. One notable challenge was the changes over the years in the formulation of the collection, both in terms of technology and personnel. With every change, there are discrepancies in how the collection is managed, accessed, and used. For example, UIUC Library recently transitioned our ILS from Voyager to Alma. In the migration process, the metadata may have some gaps and limitations, which could have affected the bibliographic records that we exported.

Moving Forward

As the project comes to an end, several steps can be taken to make the most of its deliverables for future initiatives. The work completed during the project provides valuable insights into assessing our international government information collection and its values for the broader community. The documentation can be helpful to inform and guide future projects, providing a roadmap that can be used to streamline similar ideas and simplify processes.

Currently, the Government Information Team at UIUC, which is comprised of graduate assistants and our Government

Subjects	Counts
Nutrition; Rural development; Agriculture	24312
India--Politics and government--1947--Periodicals	2334
Gazettes--France.; Law--France--Periodicals.; France--Politics and government--1870-1940--Periodicals	2030
Great Britain--Politics and government--1945--Periodicals	1858
Brazil--Congresso Nacional.--Câmara dos Deputados	1456
India--Politics and government--1947-; Politics and government	1441
Canada--Politics and government--1867--Periodicals.	1355
Netherlands--Politics and government.	1116
Chemins de fer--Recherche.; Railroad engineering--Research	913
Great Britain--Politics and government--1945--Periodicals.	817
Africa, North--Economic conditions--Statistics.	769
Law--Canada.; Gazettes--Canada.; Canada--Politics and government--Periodicals.	764
India--Census, 1961.	760
Finland--Politics and government	716
India--Commerce.; Aussenwirtschaft; Indien	646
Muhammad,--Prophet,--632.; British Library.--Oriental and India Office Collections.; Muslims--Biography--Early works to 1800	558
Germany.--Bundestag--Periodicals.; Germany--Politics and government--1990--Periodicals.	542
European communities--Periodicals.; Europees recht	488
British Library.--Oriental and India Office Collections.; Islamic civilization--Early works to 1800.; Civilization, Arab--Early works to 1800	463
Germany--Politics and government--1871-1918--Periodicals.; Germany--Politics and government--1918-1933--Periodicals.	459

Figure 5. This shows the counts of publications according to Library of Congress Subject Headings

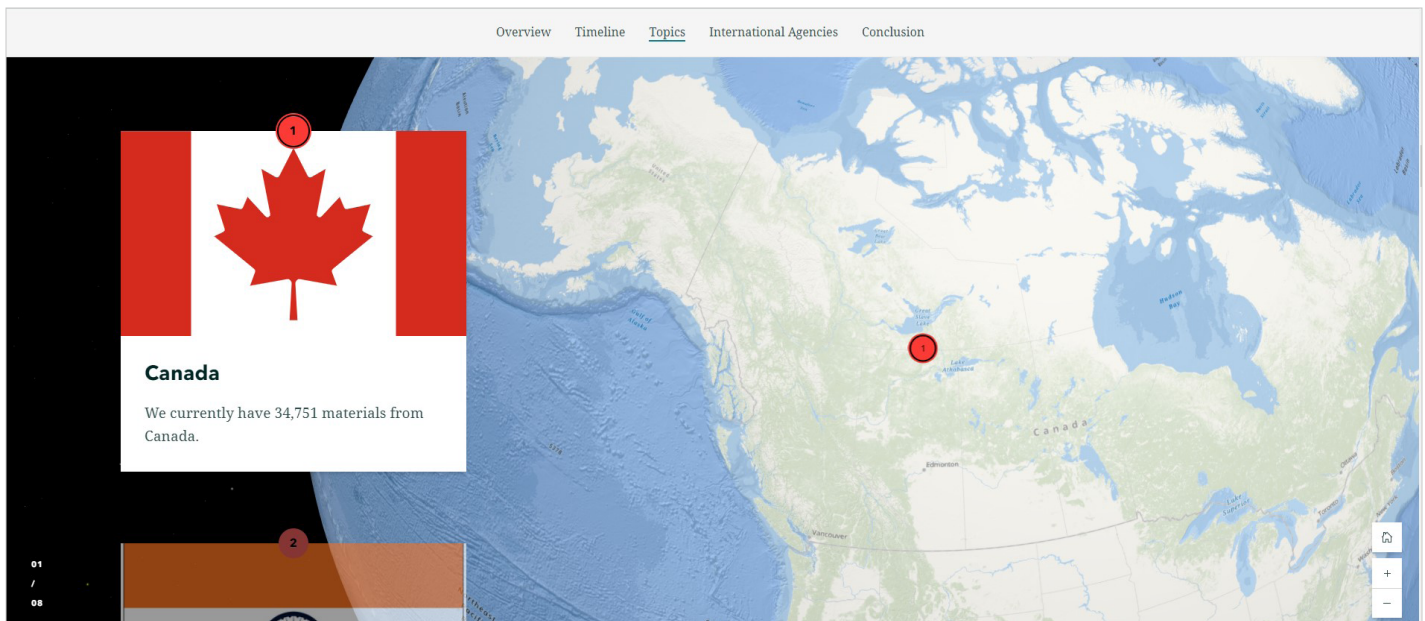


Figure 6. An example of representing publication countries with ArcGIS StoryMaps

Information Librarian, is also exploring ways to visualize other parts of the collection. We hope to enhance the accessibility and comprehensibility of the information, and devise ways to present our collection in a more engaging manner.

Additionally, collaborating with other units in the libraries such as Assessment, Reference and Instruction, Circulation, etc. can provide holistic perspectives on how the international government information collection is situated within a broader conversation of library collection and resources. For example, it would be useful to look at the usage data of the collection to see how much of the collection is being utilized

and what subjects our community is interested in for more focused outreach efforts. Some of the graphics will be used in our Libguides and other public venues to promote our collection as well. These visualization efforts add value to our existing collection, and open new opportunities for research in this area.

The project can also benefit from a more in-depth qualitative analysis. Since this project mainly focused on quantitative analysis, much of our textual data remains unanalyzed. By delving deeper into our collection with qualitative analysis tools, future research activities can uncover more insights and

nuances, enriching our understanding of the collection and its subject content.

Conclusion

This project gave us the opportunity to be more intentional in the ways we build collections and offer services as part of the Government Information Team at UIUC. By understanding the breadth and gaps in our collection, we can ensure that our resources align with the needs of our users and the community. In addition, this analysis also aids our decision-making process to identify priorities of our collection for allocation of fundings and resources. Moreover, this project served as a catalyst for promoting our collection, amplifying its visibility and accessibility. The final ArcGIS StoryMap will be used as promotion materials to introduce our collection to our users and share widely with library colleagues.

Uyen Nguyen (uyennguyentu00@gmail.com) is a recent MLIS graduate from the University of Illinois at Urbana-Champaign School of Information Sciences. This paper was written as part of a project during Nguyen's Graduate Assistantship at the University of Illinois Urbana-Champaign Library, Spring 2024, under supervision of Assistant Professor Sanga Sung. She is currently the Electronic Resources and Discovery Librarian at Duke University Medical Center Library & Archives.

References

1. "Government Information Services—University of Illinois Library," accessed August 26, 2024, <https://www.library.illinois.edu/govinfo/>.
2. "Government Information Services—University of Illinois Library."
3. Tamara Munzner, *Visualization Analysis and Design* (New York: A K Peters/CRC Press, 2014), <https://doi.org/10.1201/b17511>.
4. Jannette L. Finch and Angela R. Flenner, "Using Data Visualization to Examine an Academic Library Collection," *College & Research Libraries* 77, no. 6 (2017): 765, <https://doi.org/10.5860/crl.77.6.765>.
5. Ilka Datig and Paul Whiting, "Telling Your Library Story: Tableau Public for Data Visualization," *Library Hi Tech News* 35, no. 4 (January 1, 2018): 6–8, <https://doi.org/10.1108/LHTN-02-2018-0008>.
6. Mark Eaton, "Seeing Library Data: A Prototype Data Visualization Application for Librarians," *Journal of Web Librarianship* 11, no. 1 (January 2, 2017): 69–78, <https://doi.org/10.1080/19322909.2016.1239236>.
7. Kathryn M. Wissel and Lisa DeLuca, "Telling the Story of a Collection with Visualizations: A Case Study," *Collection Management* 43, no. 4 (October 2, 2018): 264–75, <https://doi.org/10.1080/01462679.2018.1524319>.